ChinaXiv合作期刊 图书情报工作
第65卷 第22期 2021年11月

人工智能时代文本与数据挖掘合理使用规则设计研究*

■ 吴高[1,2]　黄晓斌[2]
[1] 广西师范大学法学院　桂林 541004　[2] 中山大学信息管理学院　广州 510006

摘　要：[目的/意义]在对美国、英国、欧盟、日本、德国等地区或国家对 TDM 行为的立法或司法应对进行比较分析基础上，提出我国 TDM 合理使用规则的设计思路。[方法/过程]首先分析人工智能时代文本与数据挖掘的技术特征及其挑战，介绍出版商和图书馆有关 TDM 合理使用的立场分歧，之后在对代表性国家有关 TDM 合理使用法律应对机制进行比较分析基础上，从 TDM 例外的主体、客体、目的、行为及其他条件等方面，论证我国 TDM 合理使用规则的具体内容。[结果/结论]任何机构或个人基于科学研究或其他合理目的对任何作品以复制、提取、改编或有限传播方式进行文本和数据挖掘构成合理使用，但利用非合法来源作品须支付合理的使用费。

关键词：人工智能　文本与数据挖掘　TDM　合理使用　例外
分类号：G250
DOI：10.13266/j.issn.0252－3116.2021.22.001

人工智能被看作是21世纪三大尖端技术（基因工程、纳米科学、人工智能）之一，许多国家出台规划将人工智能上升为国家战略。当前人类已经从信息时代迈进了数据驱动的“智能时代”，人工智能需要通过大数据学习来建立其智能。文本与数据挖掘（Text and Data Mining，TDM）是众多交叉学科的重要支撑技术，是人工智能、区块链、云服务等技术应用的基础，不仅对加速科研发现、推动技术创新、创新商业模式、促进经济发展具有重要价值，而且在公共文化行业大有可为，为图书馆等机构基于 TDM 提供个性化智能服务提供坚强技术保证。然而由于运用 TDM 技术很可能会对受著作权保护的作品或数据库内容进行复制、提取、重构和传播，该行为是否属于合法行为一直是理论界和实务界热议的焦点。大部分国家和地区还未有效建立 TDM 合理使用规则，只有欧盟、英国、日本、美国等少数地区与国家的立法和司法给予了相应回应。根据 CNKI 数据库统计，国内有关 TDM 著作权问题的文献约有20余篇，并且现有理论研究更多注重于对部分国家立法经验的介绍，系统论证我国 TDM 合理使用规则具体内容的文献较为少见。新技术的发展必然会打破著作权法的传统利益平衡机制，如何重构新技术环境下著作权利益平衡，构建适应我国国情的 TDM 合理使用新型规则对于促进我国科研创新和经济发展具有重要意义。故本文在全面比较美国、英国、欧盟、日本、德国等国家或地区法律应对机制基础上，详细论了我国 TDM 合理使用规则设计内容及其理由。

1　人工智能时代文本与数据挖掘技术特征及其挑战

1.1　人工智能时代 TDM 的技术特征

TDM 是一般指从大量文本或数据中提取数据、组织信息和发现知识的过程。2014年英国修订的著作权法将 TDM 定义为利用自动分析技术对文本和数据进行分析以获取模式、趋势及其他有用信息[1]。2019年3月26日欧盟议会通过的《数字单一市场版权指令》（Directive on Copyright in the Digital Singles Market）[2]认为 TDM 是指任何旨在对数字格式的文本和数据进行分析的自动分析技术，以便生成包括但不限于模型、趋势、相关性等在内的信息。TDM 具有如下技术特征：

1.1.1　主体的特殊性

TDM 实施不仅需要进行挖掘算法开发，还需要经过数据准备与管理、数据预处理与转化、算法调试与应用、结果分析与呈现等复杂流程，个人一般不具备对海量文本和数据进行复制、提取、处理、对比、分析的条件

* 本文系广西文科中心人工智能专项重点项目“人工智能良性发展的法律规制问题研究”（项目编号：ZN2018001）研究成果之一。
作者简介：吴高，研究馆员，博士研究生，硕士生导师，E-mail：25863158@qq.com；黄晓斌，教授，博士，博士生导师。
收稿日期：2021－04－16　修回日期：2021－08－26　本文起止页码：3－13　本文责任编辑：易飞

和能力,实施者一般为具备一定技术和物质条件的组织机构(例如图书馆、大学、企业或其他研究机构等)。

1.1.2 对象的广泛性

包括文字、图片、视频、声音、数据等在内的任何数字形式材料都可成为数据挖掘对象。考虑到与研究主题相关的材料获取越多,则通过挖掘技术分析自动得出的结论更具通用性与准确性,因此 TDM 采用"样本 = 总体"的全数据模式来最大范围获取全部相关材料。

1.1.3 过程的转化性

过程的转化性既是指数据格式的转化,所获取的各类型数据(包括非结构化数据、半结构化数据和结构化数据)须转化成能够被计算机处理的结构化数据,又是指数据挖掘过程中不可避免会对受著作权保护的文本数据进行复制,但是输出结果并不直接呈现原文内容,而是将其作为分析背景资料以发现模式、趋势及规律等信息。

1.1.4 结果的价值性

作为发掘潜在价值的重要工具,TDM 在商业、教学、科研及社会管理中的应用蕴含着巨大的经济价值和社会机会,尤其是对于医学、制药、金融以及其他需要进行市场分析的领域,如提高科研效率,发掘隐藏信息并开发新知识,改进研究过程和基础,探索新的领域,等等。

1.2 利用 TDM 所面临的侵权风险

TDM 所挖掘的文本或数据,包括任何不受著作权保护或受著作权法保护的材料。TDM 实施流程往往需要对材料进行复制处理,但不同的 TDM 技术复制的方式或数量具有很大差异性,大多时候 TDM 分析的前提是重复复制整个作品[3],但有时候 TDM 技术仅对目标文本进行"逐个""单独"处理,每次只复制单个或少量的单词或数据,而且并不保留或固定抓取副本,只是计算单词或数据出现次数,此时 TDM 行为一般不构成著作权法意义上的复制行为。本文主要探讨对目标文本进行大量复制的 TDM 技术的实施所导致的侵权风险。具体包括:

1.2.1 可能侵犯著作权人的复制权

我国于 2020 年 11 月修订的《著作权法》将复制权定义为:以印刷、复制、拓印、录音、录像、翻录、翻拍、数字化等方式将作品制作一份或者多份的权利。我国学理界一般认为,构成著作权法意义上的"复制行为"须满足两个条件,即应在有形物质载体上再现作品,而且作品应"固定"在有形载体上[4]。TDM 实施流程一般包括信息抽取、语义分析、关系计算及知识发现 4 个步骤。在对材料进行挖掘处理过程中,往往会涉及到大量的版权作品,不管是将这些数据读入系统或者是进行格式转换和数据分析,均涉及到受著作权人控制的复制行为。由于大量孤儿作品存在以及系统无法有效识别目标对象的权利状态,若行为人未获得权利人授权或不符合相关侵权豁免条件,则 TDM 行为很可能侵犯著作权人的复制权。

1.2.2 可能侵犯著作权人的数据库权

对于具有独创性的数据库由复制权予以保护,对于不具有独创性的数据库是否提供著作权保护,各国意见并不一致。为更好促进数据库产业发展,欧盟率先颁布《数据库保护指令》(1996)[5],建立了新的特殊权利——数据库权,随后德国将非独创性数据库纳入邻接权保护范围,英国则对数据库进行单独立法保护。在美国由于广大公众对"信息自由"损害和权利垄断的担忧,美国并没有建立新的特殊权利制度,转而采取反不正当竞争法进行保护。在欧盟地区只要制作者对数据库相关设施设备进行了实质性投入,则可获得数据库权的特殊保护。根据《数据库保护指令》第 7 条可知,数据库权的主要内容是制作者有权禁止抽取(extraction)(将数据库的全部或实质性部分以任何方式或手段永久性或暂时性地转移到其他介质之中)和再利用(re-utilization)(通过复制、出租、网络或其他传输方式,以任何形式向公众提供数据库的全部或实质性部分)数据库的全部或实质性部分。在 TDM 过程中,往往不可避免地会涉及到抽取行为,若行为人未获得权利人授权,又无其他豁免事由,则可能侵犯著作权人的数据库权。须指出,由于我国并未规定数据库权,故针对非独创性数据库的 TDM 侵权风险较低。

1.2.3 可能侵犯著作权人的其他权利

从大的权利类型来看,TDM 行为人还可能侵犯著作权人的演绎权和传播权。我国著作权法未直接采用演绎权概念而是将其分解为翻译权、改编权、汇编权等类型。从权利内容来看,TDM 行为人很可能侵犯著作权人的改编权,因为 TDM 技术应用需要对目标文本进行识别和转码,转码行为是"改变、编排目标对象的表达形式,从而形成新的研究样本"[6],TDM 过程中的"转码行为"与我国著作权法意义上的"改编行为"在很大程度上具有同质性。此外,作为受技术发展影响最为明显的一项权利,传播权是继复制权、演绎权后出现的著作权经济权利的统称,包括表演权、出租权、广播权、展览权和信息网络传播权等类型。我国著作权法没有"传播权"概念,而是通过增加"信息网络传播

吴高,黄晓斌.人工智能时代文本与数据挖掘合理使用规则设计研究[J].图书情报工作,2021,65(22):3-13.

权"来回应新技术的发展。数据挖掘最终分析结果既可能以简单结论方式呈现,也可能以长篇幅的商业或学术报告方式进行呈现,不管哪种呈现形式均可能涉及原作品表达或数据库权利内容,若 TDM 行为通过线下或线上方式向公众公开该内容就很可能侵犯著作权人的传播权。

1.3 利用 TDM 所面临的法律困境

由于 TDM 属于新兴技术,尽管欧盟、日本、英国等少数国家及时修订法律将 TDM 纳入著作权合理使用范围,但大多数国家还未来得及针对 TDM 技术做出制度安排回应,其现有的著作权合理使用规定仅能涵盖很少量满足条件的 TDM 行为,绝大部分 TDM 行为仍面临很大的法律适用困境。例如我国于 2020 年 11 月出台了新《著作权法》,但缺乏 TDM 专项条款,TDM 行为无法满足该法第 24 条第 1 款第 1 项(个人使用)、第 2 项(适当引用)、第 6 项(教学科研)、第 8 项(文化机构)等合理使用规定。即便就已出台了相应规则的国家而言,由于受到个案判断或适用条件限制,TDM 行为是否构成合理使用也面临法律上的不确定性。目前 TDM 合理使用机制主要分为两种类型,即以美国为代表的灵活例外制度和以欧盟为代表的法定例外制度,但这两类制度也面临相应困境。

1.3.1 灵活例外制度面临个案判断困境

美国版权成文法中并没有有关 TDM 的明确规定,而是通过适用第 107 条合理使用"四要素"来对 TDM 行为是否构成合理使用进行综合判断,其中最典型的案件是 Google 数字图书馆案[7]和 Hathitrust 数字图书馆案[8]。法院最终认定 Google 和 Hathitrust 所实施的 TDM 行为具备高度的转换性,构成合理使用。虽然美国灵活例外很大程度上推动了 TDM 技术的应用和发展,但该制度也存在较大缺陷,由于依靠司法个案进行合理性判断,无法针对 TDM 行为向公众提供稳定的构成合理使用的法律预期,也不利于人工智能产业发展,如在美国 2018 年的"TVEyes 案"[9]中,第二巡回法院推翻一审法院部分判决结论,认为 TYEyes 公司为公众提供电视广播内容视频片段剪辑(不超过 10 分钟)的搜索和浏览服务,很可能构成原作品的替代品,不构成合理使用。

1.3.2 法定例外制度面临条件过严困境

欧盟传统著作权法对 TDM 缺乏明确的适用空间,如欧盟《信息社会版权指令》(2001)[10]第 5 条第 1 款规定的临时复制例外、第 5 条第 2 款和第 3 款规定的基于非商业性利用目的非强制性的个人使用例外和科学研究例外以及《数据库保护指令》(1996)第 9 条规定的基于非商业性利用目的的非强制性的个人使用例外和科学研究例外,无法有效满足大多数 TDM 行为。欧盟《数字单一市场版权指令》(2019)提出了基于科研目的以及文本与数据挖掘的两种 TDM 例外,虽然这两项例外能促进 TDM 行业发展,但基于科研目的的例外受限于科研目的,适用范围较窄;而基于文本与数据挖掘的例外,则附加了权利人的权利声明保留要求,这使 TDM 行为面临潜在的侵权风险[11]。

2 有关文本与数据挖掘合理使用立场分歧

2.1 出版商有关 TDM 合理使用的立场

随着 TDM 技术不断发展,其法律确定性问题日益突出。纵观著作权法发展历史可知,权利人的"权利束"不断丰富和扩张,以适应新传播技术的发展。针对新出现 TDM 技术可能给权利人利益造成负面影响的情况,出版商普遍希望通过立法修订来赋予权利人(出版商)新的权利类型以抗衡新技术的应用,这种新的权利就是挖掘权(the right mine)。国际出版商针对用户 TDM 合理使用的立场具体包括:

2.1.1 主张通过许可协议模式达成 TDM 最佳实践

由于大多数国家或地区并未专门制定有关 TDM 的法律条款,多数出版商主张通过许可协议来解决 TDM 面临的法律确定性问题,因为通过许可协议能细致规定 TDM 行为,而"非商业研究"概念模糊不清,而且 TDM 例外条款会破坏对出版商持续投资出版高质量内容的激励机制,同时认为目前也没有确实证据表明缺乏 TDM 例外制度会导致经济或竞争落后。出版商在不同场合不断发表自己的主张,如在 2013 年欧盟就"欧洲许可证"举办了利益相关者对话,与会出版商建议就 TDM 问题采用多方合作的市场机制,承诺为非商业研究人员开展 TDM 提供相应便利,但前提是基于许可协议条款[12]。2014 年欧洲出版商协会明确反对在著作权法中引入 TDM 合理使用条款,认为通过许可协议就能满足 TDM 实践需求并降低侵权和滥用风险。国际科技医药出版商联盟(International Association of Scientific, Technical & Medical Publishers, STM)曾在 2015 年发表意见认为 TDM 法律确定性可通过许可协议实现,TDM 例外条款会破坏确保生产高质量内容的投资激励机制[13]。

2016 年 11 月 STM 针对公布的《数字单一市场版权指令(草案)》认为,该草案指令规定的例外条款会产生以下意想不到的后果:①引入公私伙伴关系的概

念可以为商业机构可能滥用 TDM 例外提供便利。若研究人员部分资金来源于商业机构，其目标是创造商业价值，则挖掘应遵循许可协议条款；或者雇佣非营利或学术机构为商业机构进行文本挖掘，必须先获得授权。②TDM 例外适用的关键是要合法获取相关内容。必须明确指出，挖掘非法内容或通过后期再利用方式挖掘不同著作权例外下的副本都是不允许的。这种例外不应允许非客户通过文献传递或作为著作权例外的受益人方式获得内容后进行挖掘。③对于出版商平台上的操作系统来说，很难区分真正的文本挖掘者和试图通过大量非法复制作品来侵犯版权的人。他们都通过自动方式大量下载资料，当前指令中关于技术保护的用词模糊不清，造成混乱，并使出版商内容面临风险。④只要提取过程一完成，作品或主题材料的副本就应立即删除。STM 还认为美国开展 TDM 须根据许可协议进行，只有在没有市场替代或给权利人造成市场损害的情况下，才可能被视为合理使用；英国和法国的版权例外都是科学研究，比指令草案更严格要求非商业性[14]。

2.1.2 主张利用出版商提供的 API 及平台实施 TDM

许多出版商主张通过应用专用程序接口（Application Programming Interface, API）或挖掘平台来提供 TDM 服务；针对跨平台许可难题，数千家出版商与 CrossRef 公司于 2014 年 5 月联合推出了 CrossRef TDM 服务[15]，以便提供通用的 API 和许可协议框架。出版商要求用户利用 API 平台开展 TDM 的依据主要有：①确保系统平台运行。若允许用户采用任意第三方软件在其数据库平台上进行数据批量抽取或下载，将会对其平台运行造成巨大压力或可能导致系统崩溃，影响用户的正常使用，加大了出版商违约的侵权风险。②具有相关法律依据。如英国 2014 年修订的著作权法例外条款相关解释文件指出，出版商可采取合理措施（如合理限制下载速度、控制特定时间内用户访问量等，Elsevier 甚至认为合理措施包括要求通过专门的 API 实施文本挖掘[16]）来维护网络安全和稳定性，只要这些措施不阻止或不合理限制任何研究人员从例外中受益的能力[17]。③区分界定两类行为。出版商对两类行为即“挖掘”和“阅读”的内涵进行严格区分，认为这是两种具有根本性区别的信息活动，需制定不同的许可协议使用条款来确定 TDM 的权利范围。

一些出版商针对其平台提供 API 服务设置了较多的限制条件，以 Elsevier 的 TDM 政策为例：由其 TDM 注册登记表可知，Elsevier 开展的 API 服务对订阅用户实施 TDM 作出了比较严格的限制[18]，具体表现为：①限制挖掘范围，只针对 XML 文件，使用图像等信息时应与相关权利人协商以获得重新利用的授权；②限制挖掘过程，明确禁止使用任何非 API 的自动下载设备或软件来获取其网站内容，并严格限制了访问数量和访问频率；③限制挖掘结果，尽管允许研究人员将挖掘结果用于存储于机构知识库中或发表论文，但禁止创建、删改、翻译任何基于语料库的衍生作品，以避免对其产品和服务产生竞争性威胁。Springer 的 TDM 政策也具有代表性。Springer 公司与美国著作权结算中心（Copyright Clearance Center, CCC）合作提供跨平台的 TDM 服务[19]，尽管 Springer 的 TDM 政策相对宽松，允许订阅机构的研究人员使用 PubMed、Web of Science 或 Springer Nature 的元数据 API 等工具来实施 TDM，并且可直接从其平台上下载全文数据内容，不强制要求注册 API 密钥，但限制下载速率，每秒 1 次请求（若注册申请 API 密钥，下载速率为每分钟 150 个请求），并要求研究人员须采取合理措施确保数据安全，如数据存储在内部安全服务器，禁止第三方访问，只能在 TDM 项目期间使用等[20]。

2.2 图书馆有关 TDM 合理使用的立场

TDM 关系到社会公众能否享有自由获取知识的权利，面对权利人针对 TDM 技术引发的著作权利益冲突所发起的紧逼和“围剿”措施，作为公共利益代言人的重要团体，图书馆界主动出击，发出自己的声音，提出“有阅读权就有挖掘权”的重要主张，在为公众积极争取 TDM 权利方面开展了一系列工作，尤其是着力解决 TDM 所面临的法律不确定问题，旗帜鲜明地阐明了国际图书馆界立场，获得了广大消费者组织的认同。

2.2.1 主张尽快完善 TDM 合理使用规则

国际图联（International Federation of Library Associations, IFLA）于 2013 年发布《关于文本与数据挖掘的声明》[21]指出，作为促进学习和创造新形式的重要工具，TDM 的法律确定性只能通过（法定）例外来实现，许可协议不能作为 TDM 的解决方案，并认为信息应该不受限制地进行利用，这对社区的教育和文化福利的增进至关重要。IFLA 认为，若缺乏 TDM 例外，研究者在开展重要研究和数据驱动的创新过程中就会面临由于法律不确定性而带来的风险。欧洲研究图书馆协会（the Association of Europe Research Libraries, LIBER）发布了《数字时代知识发现海牙宣言》[22]，提出用户享有隐私权、信息权及挖掘权，认为政策制定者应明晰法律内容，明确阅读权包括“挖掘权”，确保内容挖掘不侵

犯著作权及邻接权,并建议各大学、研究资助者、研究机构和商业机构制定鼓励采用内容挖掘研究方法的政策。

图书馆界呼吁从法律制度上完善 TDM 合理使用规则,主要缘由包括:①通过许可协议规制 TDM 具有很大弊端。在许可协议谈判过程中,由于出版商(权利人利益代表)和图书馆(公共利益代表)地位不平等,出版商往往会利用垄断优势地位,施加单边意志,通过保留修改条款权利或签订霸王条款,不合理限制用户实施 TDM 行为,协议长期有效性也难以保证。此外,由于 TDM 所利用的信息的异常丰富,已经远远超出了被许可使用的研究数据库,这也使许可使用方案变得不切实际。②制定 TDM 例外条款有利于提升国家国际竞争力。一些国家已经将 TDM 合理使用制度纳入其著作权法律体系当中,例如美国和加拿大基于合理使用框架体系规制 TDM 行为,欧盟、英国、日本则制定了专门的 TDM 合理使用条款。这些国家通过修订法律确保了 TDM 高效实施,有助于其科研人员更好把握科研创新的主动权,从而促进技术创新和经济发展。

2.2.2 反对限制使用 TDM 相关工具软件

针对大多数出版商要求用户使用专门的 API 或挖掘平台来限制 TDM 的要求,图书馆等信息服务机构则持反对态度,主要理由如下:①API 平台本身存在较多弊端。LIBER 等组织认为 Elsevier 的 API 注册条款过于严格,如只允许挖掘文本,将图片、图表、视频等排除在外,明确禁止使用机器人、蜘蛛等自动化程序,对原文片段的引用不得超过 200 个字符等。②仅允许利用 API 平台容易产生垄断。不允许科研人员利用自己或第三方的研发工具,不利于维护学术自由,推动科学研究,提高科研效率,并且侵犯了科研人员隐私权。③API 平台技术限制侵犯了读者合法阅读的权利。信息服务机构认为,阅读权包括挖掘的权利,既然机构投入大量经费购入相应数据库资源,就应该为读者提供不受限制的挖掘权[23]。

3 域外有关文本与数据挖掘合理使用的法律应对

3.1 美国司法

针对 TDM 行为面临的版权困境问题,美国并未通过单独立法予以解决,而是通过将原则性合理使用条款应用到司法判例实践中,最终以司法个案形式确认了 TDM 行为的合法性。2003 年以来美国法院在多个判决中认定 TDM 涉及的复制行为属于合理使用[24],如“Kelly 诉 Arriba Soft”案[25](2003)、“Field 诉 Google”案[26](2006)、“Perfect 10 诉亚马逊”案[27](2007)、“A. V. 诉 iParadigms”案[28](2009)、“作家协会诉 Google”案[29](2011)、“Fox 诉 TVEyes”案[30](2014)、“White 诉 West”案[31](2014)、“作家协会诉 HathiTrust”案[32](2014)等,其中以“作家协会诉 Google”案和“作家协会诉 HathiTrust”案最具代表性。上述案例中,美国法院借助版权法第 107 条有关合理使用的原则性条款,尤其是引入“转换性使用”(transformative use)理论,对 TDM 行为进行合理使用“四要素”分析,从而推理出 TDM 行为属于合理使用。“转换性使用”概念最早由 Leval 法官于 1990 年提出,系指以不同目的或不同方式对作品进行创新性使用[33]。1994 年在 Campbell v. Acuff-Rose Music 案件中转换性使用首次在司法领域被认定为合理使用[34]。之后,P. Samuelson 在上述案件基础上,将转换性使用细分为 3 种类型:①在评论时利用新的表达方式对原有作品进行转换;②适当改变原作品含义或所传递的信息,从而在原作品基础上增加了新的内容;③基于与原有目的相区别的其他目的使用原作品[35]。

在“作家协会诉 Google”案件中,Google 公司开发“Google Books 项目”,即将美国大学图书馆提供的纸本图书进行全文数字化扫描,利用 TDM 核心技术开发搜索和片段浏览功能,向公众提供高效且有创新性的数字化检索服务。具体而言,Google 公司将扫描后的图书每页分为 8 个部分,公众输入检索词后则呈现少量含有检索词的片段,这实际上就是 TDM 技术的应用。2005 年作家协会提起对 Google 的版权侵权诉讼,2009 年 11 月双方达成和解协议,Google 公司被允许以默示许可方式利用版权作品,但后被法院否决。2011 年 11 月纽约南区地方法院对该案作出初审判决,认为“Google Books 项目”具有很强的转化性目的——为公众提供完善且高效的文字搜索服务,属于合理使用。2015 年 12 月美国第二巡回法院维持原判,并补充到:“具有商业盈利不能成为构成否定合理使用的绝对标准;且片段检索模式不对版权市场构成替代效果,不会减损权利人的实质利益”[36]。在“作家协会诉 HathiTrust”案件中,Hathitrust 数字图书馆允许公众使用特定词语在其机构库中的全部数字副本中进行搜索,但搜索结果只显示搜索关键词所在的页码以及在每页当中显示该词语的频率。初审法院运用合理使用“四要素”判定方法,认为 Hathitrust 数字图书馆所提供

的检索服务不属于单纯的使用作品,属于衍生出新的学术研究方法与路径,具有较强的“转化性目的”[37]。美国第二巡回法院亦认为 Hathitrust 数字图书馆实施 TDM 行为构成合理使用[38]。由上可知,美国法院从司法层面最大程度确认了 TDM 行为的合法性,属于“无条件例外”模式,但该模式适合判例法国家,要求法官具备充分的司法经验,需要进行个案判断,效率较低且成本较高。

3.2 英国立法

2010 年英国政府启动了对版权制度的全面改革。2011 年 5 月 I. Hargreaves 教授受英国政府委托发表报告《数字化机遇:知识产权和发展回顾》[39],针对英国现有知识产权立法提出 10 项建议,其中之一就是建议政府引入非商业性研究的 TDM 行为例外。2014 年英国颁布《版权和表演权(研究、教育、图书馆和档案馆)条例》[40],在《版权、外观设计和专利法》(1988)[41]中第 29A 条(基于非商业性研究的文本与数据挖掘之复制)引入了 TDM 版权例外。该条款主要内容包括:①合法获取作品的人复制版权作品不构成侵权,只要满足以下条件:其一,复制是为了对作品任何内容进行计算分析,且唯一目的是出于非商业性的研究;其二,复制件附有充分的作者信息标示(除非因实际原因或其他原因不可行)。②若未经版权所有人允许,将复制件转让给其他人或将复制件用于非商业研究以外的目的,构成侵权。③上述复制件若被出售或出租或者提供出售或出租,则该复制构成侵权。④文本与数据挖掘例外优先于合同约定效力。

由前可知,英国版权法对 TDM 例外适用规定了 3 个前提条件:①被复制的作品须为通过合法途径所获取。这意味着科研人员只有其本人或所在机构已经购买相关资源的前提下才能对版权作品进行数据挖掘和信息分析,否则构成侵权。②复制目的须为非商业性研究目的。这意味着任何商业性的 TDM 所涉及的复制行为必须获得作品版权人的许可,否则构成侵权。③复制需要充分标明作品的作者信息。这意味着使用人须尊重版权人人身权,未充分标示作者信息构成侵权,除非因实际原因或其他原因不可行。英国版权法针对 TDM 例外也规定了相应的限制条件,即禁止将复制件转让给其他人,或将复制件用于非商业研究以外的目的,或复制件被用于“出租或出售”。值得肯定的是,英国版权法规定了 TDM 法定例外优先于合同约定效力,即任何旨在限制或阻止 TDM 所涉及复制的合同条款不具有执行力。

3.3 欧盟立法

欧盟于 1996 年和 2001 年分别颁布的《数据库保护指令》和《信息社会版权指令》由于其著作权例外条款采用封闭式立法,无法适用于 TDM,影响了 TDM 技术的发展和应用。为弥补这种缺陷并进一步推动欧盟版权法的一体化进程,欧盟于 2016 年发布《数字单一市场版权指令提案》,明确规定了基于科研目的的 TDM 版权例外。2019 年 4 月欧盟正式颁布修改后的《数字单一市场版权指令》,在规定“基于科研目的的 TDM 版权例外”之外,新增了“基于文本与数据挖掘目的的 TDM 例外”。指令第 3 条第 1 款明确规定,成员国应规定出于科学研究目的,科研机构和文化遗产机构可以对合法获取的作品或其他内容进行复制与提取,开展文本和数据分析。指令第 4 条规定,成员国应规定出于文本与数据挖掘目的,对合法获取的作品或其他内容进行复制与提取的行为不侵犯版权。

欧盟针对“基于科研目的的 TDM 版权例外”情形,规定了如下限制条件:①实施 TDM 的主体须是科研机构和文化遗产机构。科研机构是指大学(包括其图书馆)、研究所或其他主体,其主要目的在于进行科学研究或与科学研究相关的教学性活动。文化遗产机构是指可供公众访问的图书馆、博物馆、档案馆或影音遗产机构。②实施 TDM 的客体须是合法获取的作品或其他内容。来源合法的资源包括印刷形式和电子形式,即使用人合法获取作品或其他内容的纸质版和电子版,对纸质材料进行扫描和数字化,对电子材料进行下载等。③实施 TDM 的目的须是基于科学研究目的。欧盟并未对科学研究的性质进行明确限制,研究机构与商业公司合作开展的科研活动,仍可适用此例外。不过,此例外虽然将商业性研究目的涵盖在内,但有一定条件限制——研究机构应将所获利润继续投入其科学研究中,并且对该机构具有决定性影响的企业不能优先获取相关研究成果。④实施 TDM 的使用方式限于“复制”和“提取”。对于“改编”“翻译”“汇编”等类型是否适用此例外,欧盟并未明确说明。⑤实施 TDM 的挖掘结果应以适当的安全等级存储。欧盟要求存储达到一定的安全等级,以防止使用人滥用 TDM 例外,保障权利人的合法权益。

考虑到“基于科研目的的 TDM 版权例外”受到较多限制,无法广泛应用于科研以外领域的 TDM 应用(如复杂的商业决策、政府服务或新应用程序或科技的发展等),加之 TDM 分析行为不属于欧盟于《2001 年版权指令》明确规定属于“合理使用”的临时复制情

吴高，黄晓斌. 人工智能时代文本与数据挖掘合理使用规则设计研究[J]. 图书情报工作,2021,65(22):3－13.

形,欧盟新增了"基于文本与数据挖掘目的的 TDM 例外"情形。该例外适用要求是——只要权利人没有以适当方式明确保留对相关内容的使用,则使用者可对合法获取的内容进行复制与提取,没有主体资格(不限于科研机构)、使用目的(不限于非营利目的)的限制,以更好鼓励私人企业进行创新。

3.4 日本立法

2009 年日本修正版权法,在版权法第 47 条之七中引入了"为解析信息进行的复制等"著作权例外[42],该条一般被视为 TDM 著作权例外。依据该条规定,基于通过计算机进行信息分析(是指从众多作品及其他大量信息中,提取构成该信息的语言、声音、影像或其他要素,并进行比较、分类及其他统计分析)之目的情形,在必要限度内,可以将作品录入记录媒介或进行改编;但为了供信息分析者使用制作的数据库作品,不在此限。日本的"信息解析例外"为 TDM 行为提供了较好的合法性依据,具有进步意义。该例外条款特点如下:①适用对象不限于文本,还包括语言、影像、声音或其他要素,但专用于信息解析的数据库作品除外;②使用方法不限于某一特定数据分析方法,还将比较、分类或其他统计分析等方法涵盖在内;③使用方式不限于复制,还包括改编,也就是说依据该作品创作出的衍生作品的复制也可适用该例外。当然,也必须指出,日本的"信息解析例外"也有一定局限性,如将信息解析工具局限于计算机,过于局限于现有技术,忽略了随着新技术发展将来计算机之外其他先进设备设施应用的可能性。

3.5 德国立法

2017 年 9 月德国颁布了《著作权因应知识社会发展需要法案》[43],对著作权法进行修正,该法案重构了基于教育与科学目的对著作权人权利进行限制的体系,其中包括引入了文本与数据挖掘例外以及规定了法律许可使用和合同授权使用的关系,以试图解决德国著作权法被批评为无法适应新技术变化以及复杂的结构让外行人难以理解的问题[44]。德国著作权法第 60d 条为文本与数据挖掘条款,该条第 1 款规定:为基于科学研究对大量作为材料来源的作品进行自动分析,应允许使用者出于非商业目的以系统和自动等方式复制来源材料,并通过规范化、结构化和分类化的方法创建一个可用于分析的语料库;这些语料库可以向特定有限的合作开展科学研究的人群以及负责监测科学研究质量的独立的第三方公开。该条第 2 款规定依据第 1 款对数据库作品进行 TDM 利用构成惯常利用,若对数据库作品的非实质性内容进行利用视为对数据库的正常利用。该条第 3 款规定研究工作完成后原始资料的语料库和复制件应被允许传送给图书馆、档案馆、博物馆及教育机构进行长期保存。由上可知,德国有关 TDM 行为构成合理使用的适用条件如下:①使用主体不限定于某类特定主体(如科研机构或文化遗产机构等),并将商业性机构(从事非商业性研究)涵盖在内;②使用目的限定为非商业性的科学研究;③使用对象不限定于合法来源的作品,并将数据库作品也涵盖在内;④使用行为包括创建用于分析的语料库以及该语料库向特定科研人员或负责检测科研质量的第三方公开这两种豁免情形;⑤与 TDM 相关的资料可传送给图书馆等公共机构进行长期保存。

4 我国文本与数据挖掘合理使用规则的设计思路

2020 年 11 月颁布的新《著作权法》第 24 条增加了第十三项"法律、行政法规规定的其他情形"这项合理使用兜底法律条款,为我国 TDM 行为被认定为合理使用提供了一定的法律空间。但考虑到该项规定的模糊性、TDM 需求的广泛性以及我国成文法的传统,笔者建议在著作权法中专设 TDM 版权例外条款。借鉴国外新近立法经验,结合我国实际,建议 TDM 合理使用规则设计应主要包括如下内容:

4.1 TDM 例外的主体条件

欧盟 2019 年《数字单一市场版权指令》尽管未限定基于文本与数据挖掘目的的 TDM 例外的主体资格,但将基于科研目的 TDM 例外主体限定为科研机构和文化遗产机构。而美国、英国、日本、德国、法国等国家并未限定 TDM 合理使用行为的主体条件,商业性公司也可成为 TDM 合理使用行为的主体。笔者建议 TDM 合理使用行为的适用主体应不仅限于科研机构或公共文化机构,理由如下:①限定主体将无法充分利用 TDM 技术。随着技术不断发展,更善于应对市场变化的互联网或软件公司在 TDM 领域研究能力往往要强于研究机构或文化遗产机构,许多公共文化机构仅靠自身力量无法单独完成 TDM,需要与相关科技公司合作[45],因此若将主体限定为科研机构或公共文化机构,将不利于 TDM 技术的充分应用和发展。②限定主体将损害公民个人研究自由。将企业或个人完全排除成为 TDM 合理使用行为的主体,损害了我国宪法所保护的公民具有科学研究、文学艺术创作和其他文化活

动的自由，破坏了人们认识事物规律的所应具有的平等机会。③限定主体将损害社会公共利益。限定科研机构或公共文化机构将阻却其他主体通过公布TDM结果为公益事业作贡献的动机[46]，不利于推进商业性主体积极参与公益性活动。

4.2 TDM例外的客体条件

TDM例外的客体是否强制性要求具备合法来源？各国规定不尽相同。部分国家如美国、德国、日本等并不将被挖掘对象的来源合法性作为前置性条件，但也有部分国家如欧盟、英国、法国等将TDM例外客体条件限定为合法获取（lawful access）的作品。合法获取一般是指通过订阅图书期刊、购买数据库或遵守开放许可协议等多种途径依法获取作品的权利。欧盟《数字单一市场版权指令》序言第14条指出，合法获取应理解为包括依据开放获取政策，或根据机构与权利人之间的合同安排，或通过其他合法手段获取内容，或对在线免费获取的内容（权利人并未以适当方式对其复制权作出保留）进行访问。英国政府公开的文件明确指出，合法获取是指依法享有获取作品的权利，订阅图书期刊或数据库以及遵守合同协议等均为合法获取的途径[47]。笔者认为，考虑到判定被挖掘对象是否具有合法来源，需要较大的时间和经费成本，并且从促进TDM新技术应用和发展以及维护公共利益角度出发，建议TDM例外的客体应不限定为合法来源。对此问题，马克斯·普朗克创新与竞争研究所也建议科研机构在没有合法接触权的情况下应有权实施TDM，但使用人应支付合理的许可使用费[48]。此外，TDM合理使用规则是否应适用于未发表作品？有学者明确提出否定意见[49-50]。但笔者认为，尽管未发表作品属于作者绝对控制范围，但从促进知识传播、保障公众对作品接近视角看，法律应认可对未发表作品的合理使用[51]，尤其是对收藏了许多具有重要史料价值和研究价值的珍贵手稿、日记、往来信件、电报等资源的图书馆或档案馆而言，对这些未发表作品进行TDM属于生产性使用行为而非消费性使用行为，具有很强的公益性。综上所述，TDM例外的客体要求不应做任何限制，不限制于合法来源或已发表作品，但可要求使用人为非合法来源的作品利用支付合理的许可费。

4.3 TDM例外的目的条件

英国将TDM构成合理使用的目的限定为非营利或非商业性目的。欧盟则对TDM目的条件规定了更为宽松的要求。针对"基于科研目的的TDM版权例外"情形，欧盟将科研目的作为限定条件，但将商业性研究目的涵盖在内，不过要求研究机构应将所获利润继续投入其科学研究中；针对"基于文本与数据挖掘目的的TDM例外"情形，欧盟并未做相应的目的上的限制，以更好推动TDM技术在科研以外领域的应用和发展。美国通过合理使用四要素来判定TDM是否构成合理使用，美国第二巡回法院就Google数字图书馆案件明确指出：具有商业盈利不能成为构成否定合理使用的绝对标准。日本立法也并未明确限定TDM的目的条件。综合主要国家立法情况，笔者建议我国TDM构成合理使用的目的宜限定为"为科学研究和其他合理目的"，理由如下：①与非商业性目的相比，科学研究所涵盖范围更广，包括非商业性研究和商业性研究。科学研究具有公共利益的本质属性，TDM所产生的新作品并不会对原作品构成实质性替代，TDM行为本身并不会损害权利人利益，商业性科学研究具有存在的价值。②增加"其他合理目的"表述，为TDM在其他领域应用以更好实现公共利益提供保障。TDM作为一项新的数据分析技术，在科学研究以外的其他领域，如商业重大决策、政府服务（如公共健康风险评估与决策、食品安全监管与控制等[52]）、新技术或程序开发等领域都有重要应用价值。

4.4 TDM例外的行为条件

TDM的利用行为一般包括"复制""提取"（"改编"）"传播"等方式。各国针对TDM例外的适用行为规定具有差异性。如英国规定TDM例外适用行为限于复制，欧盟将其限于复制和提取，日本规定限于复制和改编。德国则规定复制和传播，但传播限于两类特定对象（特定科研人员或检测科研质量的第三方）。美国比较特殊，由于其暂时没有针对TDM例外行为的专门立法，但通过相关司法案例可知，美国法院一般将TDM例外的行为限定于复制和有限传播（不构成对原作品的实质性替代）。如在Google数字图书馆案件中，法院认为Google对原作品的片段式呈现（传播的表现形式）不足以取代原作品，构成合理使用；但在TVEyes案件中，2018年第二巡回法院推翻了2014年地方法院做出的被告对原作品的片段呈现构成合理使用的结论，认为被告为用户提供不超过十分钟的片段式浏览服务，用户可通过这些片段了解原作品的核心思想，该片段已对原作品构成了市场威胁，不属于合理使用。综上所述，考虑到复制和提取或改编是文本与数据挖掘应用中最基础和必不可少的环节，而且"传播"行为若完全禁止不符合公共利益，因此笔者建议TDM例外行为方式包括"复制""提取或改编"以及

"有限传播"(面向科研人员等特定对象进行传播或面向公众传播时传播内容不足以替代原作品)。

4.5 TDM 例外的其他条件

TDM 例外的其他条件包括以下内容:①安全保存措施。TDM 挖掘结果需要安全地存储,以供后续的数据分析或对分析结果进行验证。许多国家立法例都严格规定了 TDM 的后续存储行为,值得我国借鉴。如欧盟规定为实施 TDM 的作品复本应以适当的安全等级存储,以用于科学研究或验证研究结果。德国规定研究完成后应删除原始资料的语料库和复制品,不再向公众开放,但允许将其交由图书馆、档案馆、博物馆和教育机构长期存储。在美国 Google 数字图书馆案中,Google 重点证明其采取了严格的安全保存措施,以防止数字图书外泄和传播。②技术措施规避例外。数字时代越来越多人采取技术措施对其作品进行保护,若不允许使用人可以合法规避技术措施,那么 TDM 行为将无法实施。从全球立法来看,许多国家都规定了多项技术措施规避法定例外情形,尤其是将图书馆、档案馆等机构作为技术措施规避例外的主体,如美国、法国、德国、澳大利亚等,这在一定程度上有利于公共机构实施 TDM。但也须指出,上述法定例外还不能完全满足 TDM 的应用和发展需要,故一些地区或国家专门立法规定技术保护措施不应妨碍 TDM 的实施。如欧盟《数字化单一市场版权指令》第 3 条第 3 款规定:"权利人可采取措施确保承载作品或其他受版权保护内容的网络和数据库的安全性和完整性,但该措施不应超过实现这一目标所必需的限度",马克斯·普朗克创新与竞争研究所也建议权利人所采取的必要技术保护措施不得不必要地妨碍文本与数据挖掘[47]。就我国而言,2020 年新修订的《著作权法》仅规定了技术措施规避例外的 5 种法定情形——为课堂教学或科学研究、为盲人利益、执行公务、安全测试以及加密或反向工程研究,甚至都没有规定图书馆等机构作为合法规避技术措施的主体地位,这远不能满足文本与数据挖掘需要,故我国尤其需要规定针对 TDM 的技术措施规避例外条款。③优先于合同约定条款。为避免数据库商等权利人利用其垄断地位借助合同协议垄断条款来排除 TDM 的应用,我国有必要通过立法规定任何排除 TDM 的合同条款都属于无效约定,如英国《版权、外观设计和专利法》第 29A 条第 5 款、德国《著作权和邻接权法》第 60g 条第 1 款作了类似规定。④充分标注版权人信息。英国规定为实施 TDM 所进行的复制需要充分标明作品的作者信息,除非标注不具有可行性,如对分散的文本和数据进行标注。同时对 TDM 之目的、发生和结束之时间和地点以及 TDM 之客体等以书面或电子形式进行充分说明,除非涉及到国家安全、紧急状态等特殊情形[51]。⑤挖掘结果的使用。建议规定 TDM 结果可以免费公开,但受到一定条件限制,如公开不具备商业目的,不得未经权利人许可私下转让,不得进行相关交易行为(出售或出租,或许诺出售或出租,或为出售、出租而披露)等。

综上所述,本文建议我国 TDM 合理使用规则主要内容为:①包括科研机构、公共文化机构、商业性机构以及其他组织在内的任何机构或个人基于科学研究或其他合理目的对作品以复制、提取、改编或有限传播方式进行文本和数据挖掘,无需著作权人许可,不向其支付报酬,但应当指明作者姓名或者名称、作品名称,并且不得影响该作品的正常使用,也不得不合理地损害著作权人的合法权益。②上述文本和数据挖掘行为可以避开技术措施,但不得向他人提供避开技术措施的技术、装置或部件,不得侵犯权利人依法享有的其他权利;不禁止使用人利用非合法来源作品,但须支付合理的许可费;文本和数据挖掘结果应被安全地存储,并可基于非商业目的在不损害权利人利益的前提下向公众公开;任何旨在阻止或限制依据本条款所允许利用行为的合同协议是无效的。

参考文献:

[1] UK Intellectual Property Office. Guidance of exceptions to copyright[EB/OL]. [2021－04－13]. https://www.gov.uk/guidance/exceptions-to-copyright # text-and-data-mining-for-non-commercial-research.

[2] The European Union. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the digital single market and amending directives 96/9/EC and 2001/29/EC (Text with EEA relevance.) [EB/OL]. [2021－04－13]. https://eur-lex.europa.eu/legal-content/EN/TXT/? uri = uriserv:OJ.L_.2019.130.01.0092.01.ENG.

[3] BORGHI M, KARAPAPA S. Copyright and mass digitization: a cross-jurisdictional perspective[M]. New York:Oxford University Press,2013.

[4] 王迁. 著作权法[M]. 北京:中国人民大学出版社,2015.

[5] The European Union. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases[EB/OL]. [2021－04－13]. https://wipolex.wipo.int/zh/text/126788.

[6] 董凡,关永红. 论文本与数字挖掘技术应用的版权例外规则构建[J]. 河北法学,2019,37(9):148－160.

[7] Authors Guild v. Google, Inc., 804 F.3d 202(2nd Cir. 2015)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[8] Authors Gulid, Inc. v. Hathitrust, 755F, 3d. 87(2nd Cir. 2014)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[9] Fox News Network, LLC v. TVEyes, Inc., 883 F.3d 169(2nd Cir. 2018)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[10] The European Union. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society[EB/OL].[2021-04-13]. https://www.wipo.int/edocs/lexdocs/laws/en/eu/eu049en.pdf.
[11] ROSATI E. Copyright as an obstacle or an enabler? a European perspective on text and data mining and its role in the development of AI creativity[J]. Asia Pacific law review, 2019, 27(2): 198-217.
[12] European Commission. Licences for Europe stakeholder dialogue-frequently asked questions[EB/OL].[2021-04-13]. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_13_986.
[13] 茹丽洁,顾立平,田鹏伟. 国际出版商对文本和数据挖掘限制的正当性辨析[J]. 图书馆建设,2016(7):27-33.
[14] Ten things to know about text mining and the proposed copyright directive COM(2016) 593final[EB/OL].[2021-04-13]. https://www.stm-assoc.org/2016_11_24_2016_11_STM_Ten_things_to_know_about_TDM_and_proposed_copyright_directive.pdf.
[15] Text and data mining[EB/OL].[2021-04-13]. https://www.crossref.org/education/retrieve-metadata/rest-api/text-and-data-mining/.
[16] How does Elsevier's text mining policy work with new UK TDM law? [EB/OL].[2021-04-13]. https://www.elsevier.com/connect/how-does-elseviers-text-mining-policy-work-with-new-uk-tdm-law.
[17] Intellectual property office. Exceptions to coyright: research(2014)[EB/OL].[2021-04-13]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf.
[18] Elsevier provisions for text and data mining(TDM)[EB/OL].[2021-04-13]. https://www.elsevier.com/__data/assets/pdf_file/0012/102234/TDM-sign-up-short-form.pdf.
[19] Springer partners with copyright clearance center to power text and data mining solution[EB/OL].[2021-04-13]. http://www.copyright.com/springer-partners-with-copyright-clearance-center-to-power-text-and-data-mining-solution/.
[20] Springer Nature TDM policy[EB/OL].[2021-04-13]. https://www.springernature.com/gp/researchers/text-and-data-mining.
[21] IFLA statement on text and data mining[EB/OL].[2021-04-13]. https://www.ifla.org/publications/node/8225.
[22] LIBER. The hague declaration on knowledge discovery in the digital age[EB/OL].[2021-04-13]. https://thehaguedeclaration.com/wp-content/uploads/sites/2/2015/04/Liber_DeclarationA4_2015.pdf.
[23] LIBER calls on elsevier to withdraw TDM policy[EB/OL].[2021-04-13]. https://libereurope.eu/article/liber-calls-on-elsevier-to-withdraw-tdm-policy/
[24] 罗娇,张晓林. 支持文本与数据挖掘的著作权法律政策建议[J]. 中国图书馆学报,2018,44(3):21-34.
[25] Kelly vs. Arriba Soft, 336F. 3D 811(9th Cir. 2003)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[26] Field vs. Google, 412 F. Supp. 2d 1106(2006)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[27] Perfect 10 vs. Amazon, 508 F. 3d 1146 (9th Cir. 2007)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[28] A. V. v. iParadigms, LLC, 562 F.3d 630, 634(4th Cir. 2009)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[29] Authors Guild v. Google, 770F. Supp. 2d 666(2011)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[30] Fox News Network, LLC v. TVEyes, Inc., 43 F. Supp. 3d 379(2014)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[31] White v. West Publ'g Corp., 29 F. Supp. 3d 396(2014)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[32] Authors Guild, Inc. v. HathiTrust, 755 F. 3d 87 (2nd Cir. 2014)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[33] LEVAL P. Toward a fair use standard[J]. Harvard law review, 1990, 103(5):1105-1136.
[34] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569(1994)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[35] SAMUELSON P. Possible futures of fair use[EB/OL].[2021-04-13]. https://ssrn.com/abstract=2584180.
[36] Authors Gulid, Inc. v. Google, Inc. 804F. 3d 202(2nd Cir. 2015)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[37] Authors Guild, Inc. v. Hathitrust, 902 F. Supp. 2d 445(2012)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[38] Authors Guild v. Hathitrust, 755F.3d. 87(2nd Cir. 2014)[EB/OL].[2021-04-13]. https://www.lexisnexis.com/.
[39] HARGREAVES I. Digital opportunity: a review of intellectual property and growth[EB/OL].[2021-04-13]. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf.
[40] The copyright and rights in performances (research, education, libraries and archives) regulations[EB/OL].[2021-04-13]. https://www.legislation.gov.uk/uksi/2014/1372/contents/made.
[41] The copyright, designs and patents act 1988[EB/OL].[2021-04

－13]. https://wipolex. wipo. int/en/legislation/details/18023.
[42] 日本知识产权法[M]. 杨和义,译. 北京:北京大学出版社,2014.
[43] Act on the harmonization of the copyright law with the current requirements of the knowledge society (copyright knowledge society act)[EB/OL].[2021－04－13]. https://wipolex. wipo. int/en/legislation/details/18029.
[44] German reform on the use of copyright protected works in the fields of education and research will come into force soon[EB/OL].[2021－04－13]. http://copyrightblog. kluweriplaw. com/2018/01/15/german-reform-use-copyright-protected-works-fields-education-research-will-come-force-soon/.
[45] 王文敏,高军.人工智能时代图书馆信息分析的著作权例外规则[J].图书馆论坛,2020,40(9):60－68.
[46] 宋雅馨.文本与数据挖掘的版权例外——以欧盟版权指令修改草案为视角[J].电子知识产权,2017(6):42－51.
[47] UK intellectual property office. Exceptions to copyright: research [EB/OL].[2021－04－13]. https://www. gov. uk/guidance/exceptions-to-copyright # text-and-data-mining-for-non-commercial-research.
[48] Max Planck institute for innovation & competition. Position statement of the Max Planck institute for innovation and competition on the proposed modernisation of European copyright rules part b exceptions and limitations (Art. 3-Text and data mining)[EB/OL].[2021－04－13]. https://pure. mpg. de/rest/items/item_2383669_8/component/file_2409840/content.
[49] 唐思慧.大数据环境下文本和数据挖掘的版权例外研究——以欧盟《DSM版权指令》提案为视角[J].知识产权,2017(10):109－116.
[50] 董凡,关永红.论文本与数字挖掘技术应用的版权例外规则构建[J].河北法学,2019,37(09):148－160.
[51] 梁志文.我国著作权法上未发表作品的合理使用及其立法模式[J].法学,2008(3):101－108.
[52] 赵力.《数字时代知识发现海牙宣言》之借鉴——以内容挖掘为核心[J].图书馆,2015(9):22－26.

作者贡献说明:
吴高:负责论文初稿撰写,修改论文;
黄晓斌:负责论文修改及内容补充。

Study on the Design of Fair Use Rules for Text and Data Mining in the Age of Artificial Intelligence

Wu Gao[1,2] Huang Xiaobin[2]
[1] School of Law, Guangxi Normal University, Guilin 541004
[2] School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract: [**Purpose/significance**] Based on the comparative analysis of the legislative or judicial responses to TDM behavior in the United States, the United Kingdom, the European Union, Japan and Germany, this paper puts forward the design idea of fair use rules of TDM in China. [**Method/process**] Firstly, the technical characteristics and challenges of text and data mining in the era of artificial intelligence were analyzed, and the differences between publishers and libraries on the fair use of TDM were introduced. Then, based on the comparative analysis of the legal response mechanism of representative countries on the fair use of TDM, from the subject, object, purpose, behavior and other conditions of TDM exception, this paper demonstrated the specific content of TDM fair use rule in China. [**Result/conclusion**] Text and data mining by reproduction, extraction, adaptation or limited dissemination of any work by any institution or individual for scientific research or other reasonable purposes constitutes fair use, but reasonable royalties shall be paid for the use of works from illegal sources.

Keywords: artificial intelligence text and data mining TDM fair use exception